

# Combinatorial Entropy for Distinguishable Entities in Indistinguishable States

Robert K. Niven<sup>1,\*</sup>

<sup>1</sup>*School of Aerospace, Civil and Mechanical Engineering,  
The University of New South Wales at ADFA, Northcott Drive, Canberra, ACT, 2600, Australia.*<sup>†</sup>

The combinatorial basis of entropy by Boltzmann can be written  $H = N^{-1} \ln \mathbb{W}$ , where  $H$  is the dimensionless entropy of a system, per unit entity,  $N$  is the number of entities and  $\mathbb{W}$  is the number of ways in which a given realization of the system can occur, known as its statistical weight. Maximizing the entropy (“MaxEnt”) of a system, subject to its constraints, is then equivalent to choosing its most probable (“MaxProb”) realization. For a system of distinguishable entities and states,  $\mathbb{W}$  is given by the multinomial weight, and  $H$  asymptotically approaches the Shannon entropy. In general, however,  $\mathbb{W}$  need not be multinomial, leading to different entropy measures.

This work examines the allocation of distinguishable entities to non-degenerate or equally degenerate, indistinguishable states. The non-degenerate form converges to the Shannon entropy in some circumstances, whilst the degenerate case gives a new entropy measure, a function of a multinomial coefficient, coding parameters, and Stirling numbers of the second kind.

PACS numbers: 02.50.Cw, 02.50.Tt, 05.20.-y, 89.20.-a, 89.70.+c

Keywords: MaxEnt; combinatorics; distinguishability; occupancy; Stirling number; Bell number

## I. INTRODUCTION

Of the many interpretations of the entropy concept, the *combinatorial* (or *probabilistic*) basis of entropy was given by Boltzmann [1] and Planck [2] in the famous equation:

$$S_N = NS = k \ln \mathbb{W} \quad (1)$$

where  $S_N$  is the total thermodynamic entropy of the system,  $S$  is the entropy per unit entity,  $N$  is the number of entities,  $\mathbb{W}$  is number of ways in which a specified realization of a system can occur, known as its statistical weight, and  $k$  is the Boltzmann constant. This can be rewritten to give the dimensionless entropy [3, 4, 5, 6]:

$$H = \frac{S}{k} = \frac{1}{N} \ln \mathbb{W} \quad (2)$$

If the weight is of multinomial form, i.e.  $\mathbb{W}_{mult} = N! / \prod_{i=1}^s n_i! = \binom{N}{n_1, n_2, \dots, n_s}$ , where  $n_i$  is the number of entities in the  $i$ th state, from  $s$  such states, then in the asymptotic limits  $N \rightarrow \infty$ ,  $n_i \rightarrow \infty$ ,  $\forall i$ , using the Stirling approximation [7],  $\ln m! \approx m \ln m - m$  (or using Sanov’s theorem [8]), the entropy converges to the Shannon function [9]:

$$H_{Sh} = - \sum_{i=1}^s p_i \ln p_i \quad (3)$$

where  $p_i = n_i/N$  is the probability of the  $i$ th state. However, it must be recognised that a system may not be of multinomial weight. The best-known examples are the three distributions examined in quantum physics [10, 11, 12, 13, 14]:

- *Degenerate Maxwell-Boltzmann statistics*, in which distinguishable entities are allocated to distinguishable states, with  $g_i$  degenerate sub-states within each state (this reduces to the multinomial case for  $g_i = 1, \forall i$ );
- *Bose-Einstein statistics*, in which indistinguishable entities are allocated to distinguishable, degenerate states; and
- *Fermi-Dirac statistics*, also with indistinguishable entities allocated to distinguishable, degenerate states, but with a maximum of one entity per state;

The weights and entropy functions of these statistics are well known [e.g. 15, 16, 17]. In such cases, maximisation of the combinatorial entropy defined by (2) (“MaxEnt”), subject to the constraints on a system, *always* yields the realization of maximum probability (“MaxProb”) (or, in the non-asymptotic case, a distribution close to the maximum) [3, 4, 5, 6]. This provides a much stronger (purely probabilistic) definition of the entropy concept than that given by axiomatic or information-theoretic reasoning.

The aims of this work are (i) to review the concept of distinguishability, so often used in physics (§II), and (ii) to derive the statistical weight and entropy of a system in which distinguishable entities (balls) are allocated to *indistinguishable* states (boxes), for both non-degenerate and equally degenerate cases (§III-IV). Although this occupancy problem has a long history [7, 18, 19, 20] and is included in combinatorial classification schemes [22, 23, 24], its connection to entropy does not appear to have been examined previously.

## II. ON DISTINGUISHABILITY

The concept of *distinguishability* strongly affects the choice of statistic used for analysis. Firstly, the entities and/or states of a system might be fundamentally indistinguishable (as is currently believed in quantum

\*Electronic address: r.niven@adfa.edu.au

<sup>†</sup>Presently at Niels Bohr Institute, University of Copenhagen, Copenhagen Ø, Denmark.

physics); the statistic is thus pre-ordained. A second, more interesting case is when we *choose* whether to distinguish the entities and/or states, based on the *purpose* for which the entropy measure will be used. This is illustrated by the allocation of physicists (entities) to the seats of a bus (boxes). Four scenarios arise:

- A conference organiser is requested by physicists A, B and C for window seats, while X and Y require seats near the door; also, everyone is concerned about the likely argument between Q and T, should they be seated together. In this case, it is necessary to distinguish both the physicists and seats, leading to degenerate multinomial (Maxwell-Boltzmann) statistics (or a variant thereof, with a maximum of  $m$  physicists per seat).
- The bus company wishes to model the wear and tear on its seats. Here they have no interest in distinguishing the physicists, but need to distinguish the seats. This leads to Bose-Einstein statistics (or an intermediate variant).
- Alternatively, the conference organiser does not have any seat-specific requests, but is concerned about who will sit together. Here the physicists are distinguishable but the seats are not, leading to a new type of statistic (examined herein).
- Finally, a more disinterested observer (e.g. a traffic engineer) does not care who the passengers are, or where they sit, but needs to model whether the bus schedule is sufficient to meet demand. Here both the physicists and seats are indistinguishable.

Such considerations lead naturally to the “subjective” (or “observer-dependent”) view of the entropy concept, a viewpoint vigorously defended by Jaynes [21] [c.f. 3, 6].

### III. THE NON-DEGENERATE CASE

We now consider the number of ways in which  $N$  distinguishable balls can be allocated to  $s$  non-degenerate, indistinguishable boxes, to give the realization  $\{n_i\}$  of numbers of balls in each box (the boxes being unlabelled), as shown in Figure 1. This statistical weight can be denoted  $\mathbb{W}_{D:I} = \left\{ \left\{ \begin{smallmatrix} N \\ n_1, n_2, \dots, n_s \end{smallmatrix} \right\} \right\}$ , with  $\sum_{i=1}^s n_i = N$ . It is known [25] that the number of ways to arrange  $N$  distinguishable balls in  $k$  non-empty indistinguishable boxes (for  $k \leq s$ ) is given by the Stirling number of the second kind  $\left\{ \begin{smallmatrix} N \\ k \end{smallmatrix} \right\}$ , the first few values of which are listed in Table I. These satisfy the recurrence relation [25]:

$$\left\{ \begin{smallmatrix} N \\ k \end{smallmatrix} \right\} = \left\{ \begin{smallmatrix} N-1 \\ k-1 \end{smallmatrix} \right\} + k \left\{ \begin{smallmatrix} N-1 \\ k \end{smallmatrix} \right\}, \quad \left\{ \begin{smallmatrix} N \\ 1 \end{smallmatrix} \right\} = \left\{ \begin{smallmatrix} N \\ N \end{smallmatrix} \right\} = 1. \quad (4)$$

By combinatorial enumeration, it is readily determined that  $\left\{ \begin{smallmatrix} N \\ k \end{smallmatrix} \right\}$  does not give  $\mathbb{W}_{D:I}$ ; e.g.  $\left\{ \begin{smallmatrix} 5 \\ 3, 1, 1 \end{smallmatrix} \right\} = 10$  and  $\left\{ \begin{smallmatrix} 5 \\ 2, 2, 1 \end{smallmatrix} \right\} = 15$ ; it is their sum which gives the Stirling number  $\left\{ \begin{smallmatrix} 5 \\ 3 \end{smallmatrix} \right\} = 25$ . By definition, this gives the general result:

$$\left\{ \begin{smallmatrix} N \\ s \end{smallmatrix} \right\} = \left\{ \begin{smallmatrix} N \\ k \end{smallmatrix} \right\} = \sum_{\text{all } \{n_i\}, \text{ fixed } k} \left\{ \begin{smallmatrix} N \\ n_1, n_2, \dots, n_k, 0, \dots, 0 \end{smallmatrix} \right\} \quad (5)$$

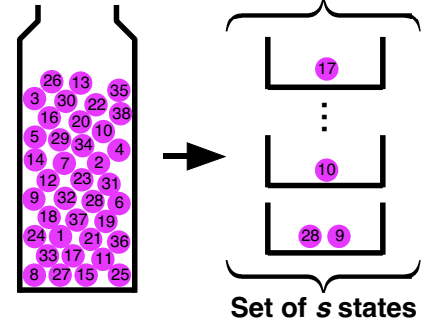


FIG. 1: Allocation of distinguishable balls to indistinguishable boxes.

	k=1	2	3	4	5	6	7
N=1	1						
2	1	1					
3	1	3	1				
4	1	7	6	1			
5	1	15	25	10	1		...
6	1	31	90	65	15	1	
7	1	63	301	350	140	21	1

TABLE I: Stirling numbers of the second kind  $\left\{ \begin{smallmatrix} N \\ k \end{smallmatrix} \right\}$ .

in which the zeroes (unfilled states) extend from  $n_{k+1}$  to  $n_s$ . Of course, some Stirling numbers  $\left\{ \begin{smallmatrix} N \\ k \end{smallmatrix} \right\}$  permit only one realization, e.g.:

$$\left\{ \begin{smallmatrix} N \\ 1 \end{smallmatrix} \right\} = \left\{ \left\{ \begin{smallmatrix} N \\ N, 0, \dots, 0 \end{smallmatrix} \right\} \right\} = 1 \quad (6)$$

$$\left\{ \begin{smallmatrix} N \\ N \end{smallmatrix} \right\} = \left\{ \left\{ \begin{smallmatrix} N \\ \underbrace{1, 1, \dots, 1}_{N \text{ times}}, 0, \dots, 0 \end{smallmatrix} \right\} \right\} = 1 \quad (7)$$

$$\left\{ \begin{smallmatrix} N \\ N-1 \end{smallmatrix} \right\} = \left\{ \left\{ \begin{smallmatrix} N \\ \underbrace{2, 1, \dots, 1}_{N-1}, 0, \dots, 0 \end{smallmatrix} \right\} \right\} = \binom{N}{2} \quad (8)$$

What can we say about  $\mathbb{W}_{D:I}$ ? Firstly, it is unaffected by any zeroes amongst the  $n_i$ , since we could arbitrarily add unfilled states to (5), without change. Secondly, as the states are unlabelled, it is meaningless to permute the  $n_i$ ; e.g.  $\left\{ \begin{smallmatrix} 5 \\ 2, 2, 1 \end{smallmatrix} \right\}$  and  $\left\{ \begin{smallmatrix} 5 \\ 2, 1, 2 \end{smallmatrix} \right\}$  refer to the *same* realization  $\{2, 2, 1\}$ . This is quite different to the multinomial weight; e.g.  $\binom{5}{2, 2, 1}$  and  $\binom{5}{2, 1, 2}$  are numerically equal, but represent different realizations  $[2, 2, 1]$  and  $[2, 1, 2]$ . The  $D : I$  statistic thus has fewer realizations than the multinomial case.

It can in fact be shown that:

$$\begin{aligned}\mathbb{W}_{D:I} &= \left\{ \left\{ n_1, n_2, \dots, n_k, 0, \dots, 0 \right\} \right\} = \binom{N}{n_1, n_2, \dots, n_k} \left( \prod_{j=1}^N \frac{1}{r_j!} \right) \\ &= \frac{N!}{\left( \prod_{i=1}^k n_i! \right) \left( \prod_{j=1}^N r_j! \right)} = \frac{N!}{\left( \prod_{i=1}^s n_i! \right) \left( \prod_{j=1}^N r_j! \right)}\end{aligned}\quad (9)$$

where  $r_j \geq 0$  is the number of occurrences of integer  $j$  in the set  $\{n_i\}$ , or its *repetitivity* (note that the zeros are not counted), whence  $\sum_{j=1}^N r_j = k$ . Proof of (9) considers the successive filling of boxes:  $\mathbb{W}_{D:I}$  must equal the number of ways to choose  $n_1$  balls from  $N$  balls, multiplied by the number of ways to choose  $n_2$  balls from  $N - n_1$  balls, and so on, for  $k$  boxes; the product must then be divided by the number of ways that each multiply-occurring integer  $j$  can occur in the set  $\{n_i\}$ , given by  $r_j!$ , to account for the indistinguishable boxes. For example,  $\left\{ \left\{ 2, 2, 1 \right\} \right\} = \binom{5}{2} \binom{3}{2} \binom{1}{1} \frac{1}{2!} = 15$ , or, since the order of filling is immaterial,  $\left\{ \left\{ 2, 2, 1 \right\} \right\} = \binom{5}{1} \binom{4}{2} \binom{2}{2} \frac{1}{2!} = 15$ . Generalising this result:

$$\mathbb{W}_{D:I} = \binom{N}{n_1} \binom{N - n_1}{n_2} \dots \binom{N - n_1 - \dots - n_{k-1}}{n_k} \left( \prod_{j=1}^N \frac{1}{r_j!} \right) \quad (10)$$

The product of binomial coefficients in (10) is simply the multinomial coefficient, giving (9); the last form in (9) can be used when  $k$  is not known in advance.  $\square$

The weight has been given previously [18, 19, 20] in the form:

$$\mathbb{W}_{D:I} = (N; r_1, r_2, \dots, r_N)' = \frac{N!}{\prod_{j=1}^N (j!)^{r_j} r_j!} \quad (11)$$

Recognising each  $j$  term in (11) as one of the  $n_j$  terms, this reduces to (9). It is, however, much less useful for the derivation of an entropy function.

Also needed is the sum of the weights, given by an *incomplete Bell number* [20]:

$$\begin{aligned}B(N, s) &= \sum_{k=1}^s \left\{ \binom{N}{k} \right\} \\ &= \sum_{k=1}^s \sum_{\substack{\text{all } \{n_i\} \\ \text{fixed } k}} \left\{ \binom{N}{n_1, n_2, \dots, n_k, 0, \dots, 0} \right\}\end{aligned}\quad (12)$$

This reduces to the usual Bell number  $B_N$  [24] for  $s = N$ .

The non-asymptotic entropy, denoted  $\#$ , can now be

calculated using (2) and (9):

$$\begin{aligned}H_{D:I}^{\#} &= \frac{1}{N} \ln \mathbb{W}_{D:I} \\ &= \frac{1}{N} \sum_{i=1}^s \left( \frac{n_i}{N} \ln N! - \ln n_i! \right) - \frac{1}{N} \sum_{j=1}^N \ln r_j!\end{aligned}\quad (13)$$

where the  $\ln N!$  term is brought inside the first sum using  $\sum_{i=1}^s n_i = N$ . Application of the “traditional” asymptotic limits  $N \rightarrow \infty$ ,  $n_i \rightarrow \infty$ ,  $\forall i$ , using the Stirling approximation, as well as the corresponding limits  $r_{j \neq \infty} = 0$ ,  $r_{\infty} = k$  then gives, for  $k \rightarrow \infty$ :

$$H_{D:I} = - \sum_{i=1}^s p_i \ln p_i - \lim_{N \rightarrow \infty} \left( \frac{1}{N} \ln k! \right) = - \sum_{i=1}^s p_i \ln p_i \quad (14)$$

$H_{D:I}$  therefore converges to the Shannon entropy (3) in the traditional limits. However, this is not the full picture, as shown by the following examples.

**Examples:** The most probable realization(s) for both multinomial and  $D : I$  statistics, determined by enumerating all realizations, are listed for several values of  $N$  in Tables II-III, for two situations: (i)  $N = s$ , and (ii)  $s = 3$ . To summarise:

- As expected, the most probable realization of the multinomial statistic, subject only to the natural constraint, corresponds in all cases to the uniform distribution  $[N/s, \dots, N/s]$ , or, if unable to achieve this (due to quantisation of the balls or boxes), to a set of equiprobable local maxima centred on the uniform distribution.
- In contrast, the  $D : I$  statistic for  $N = s$  gives a step function (“staircase”) with many unfilled boxes. The number of filled states  $k$ , the number of steps and the width of each step all increase with  $N$ , but the number of unfilled states  $(s - k)$  increases even more rapidly. This preferential bunching (“cohesion”) of the filled states is very different to their preferential spreading in the multinomial case. In consequence, applying the “traditional” asymptotic limit  $n_i \rightarrow \infty, \forall i$  (in addition to  $N \rightarrow \infty$ ) is inappropriate for  $k < s$ , since it does not account for the stepped nature of the distribution (with  $n_i \ll \infty$  in many states), and also distorts the role of the  $r_j$ . Thus for  $k < s$ , the asymptotic limit in (14) does not apply. On the other hand, for  $s = 3$  (an example of  $N \gg s$ ), the most probable realizations are again non-uniform, but all boxes are filled (except for  $N \leq 3$ ). As evident from the table, the filling will become more uniform as  $N \rightarrow \infty$ , which will be consistent with  $n_i \rightarrow \infty, \forall i$ ; this then gives the asymptotic limit of (14).

The  $D : I$  statistic thus has very different convergence properties to the multinomial, being more strongly dependent on small values of  $n_i$ ; in the limit  $N \rightarrow \infty$ , it asymptotically approaches the Shannon entropy for  $N \gg s$  and  $k = s$ . Outside of these bounds, more detailed analysis is needed to identify any asymptotic

limits; until then, the non-asymptotic entropy  $H_{D:I}^\#$  (13) must be used.

#### IV. THE EQUALLY DEGENERATE CASE

We now consider a simple degenerate form of the  $D : I$  statistic, in which each indistinguishable state contains  $g$  indistinguishable sub-states, with  $n_{im}$  entities in each sub-state, whence  $\sum_{m=1}^g n_{im} = n_i$ . The weight can be denoted:

$$\mathbb{W}_{D:I(g)} = \left\{ \left\{ \begin{matrix} N \\ n_1, n_2, \dots, n_s \end{matrix} \right\} \right\}_{(g)} = \left\{ \left\{ \begin{matrix} n_{11}, \dots, n_{s1} \\ \vdots \\ n_{1g}, \dots, n_{sg} \end{matrix} \right\} \right\} \quad (15)$$

Using the reasoning of §III, the weight of each realization is given by the weight (9) of filling of the states, multiplied by the number of ways of filling *within* each state; each component of the latter must contain a sum over the possible number of filled sub-states  $\gamma = 1, \dots, \min(n_i, g)$ . This gives, for fixed  $k$ :

$$\begin{aligned} \mathbb{W}_{D:I(g)} &= \left\{ \left\{ n_1, \dots, n_k, 0, \dots, 0 \right\} \right\}_{(g)} \\ &= \frac{N!}{\left( \prod_{i=1}^k n_i! \right) \left( \prod_{j=1}^N r_j! \right)} \\ &\quad \times \prod_{i=1}^k \sum_{\gamma=1}^{\min(g, n_i)} \sum_{\substack{\text{all } \{n_{im}\} \\ \text{fixed } \gamma}} \frac{n_i!}{\left( \prod_{m=1}^{\gamma} n_{im}! \right) \left( \prod_{\ell=1}^{n_i} r_{i\ell}! \right)} \end{aligned} \quad (16)$$

where  $r_{i\ell}$  is the repetitiveness of  $\ell$  in the set  $\{n_{im}\}$ . Using (5) and (12), this simplifies to:

$$\begin{aligned} \mathbb{W}_{D:I(g)} &= \frac{N!}{\left( \prod_{i=1}^k n_i! \right) \left( \prod_{j=1}^N r_j! \right)} \prod_{i=1}^k \sum_{\gamma=1}^{\min(g, n_i)} \left\{ \begin{matrix} n_i \\ \gamma \end{matrix} \right\} \\ &= \frac{N!}{\left( \prod_{i=1}^k n_i! \right) \left( \prod_{j=1}^N r_j! \right)} \prod_{i=1}^k B(n_i, \min(n_i, g)) \end{aligned} \quad (17)$$

For non-degenerate states  $\left\{ \begin{matrix} n_i \\ 1 \end{matrix} \right\} = 1, \forall i$ , hence (17) reduces to (9). As with the non-degenerate case, the products over  $i$  in (17) can be extended to  $s$  instead of  $k$ , since for unfilled states we can take  $\sum_{\gamma=1}^0 \left\{ \begin{matrix} 0 \\ \gamma \end{matrix} \right\} = \left\{ \begin{matrix} 0 \\ 0 \end{matrix} \right\} = 1$ , or alternatively  $B(0, 0) = 1$ .

The non-asymptotic entropy for the simple degenerate case is, from (2) and (17):

$$\begin{aligned} H_{D:I(g)}^\# &= \frac{1}{N} \ln \mathbb{W}_{D:I(g)} = \frac{1}{N} \sum_{i=1}^s \left( \frac{n_i}{N} \ln N! \right. \\ &\quad \left. - \ln n_i! + \ln \sum_{\gamma=1}^{\min(g, n_i)} \left\{ \begin{matrix} n_i \\ \gamma \end{matrix} \right\} \right) - \frac{1}{N} \sum_{j=1}^N \ln r_j! \end{aligned} \quad (18)$$

In the Stirling-approximate limits  $N \rightarrow \infty$ ,  $n_i \rightarrow \infty, \forall i$ , for which  $r_{j \neq \infty} = 0$  and  $r_\infty = k$ , we can see (e.g. from Table I) that each sum over  $\gamma$  in (18) will be dominated by its largest term  $\left\{ \begin{matrix} n_i \\ \gamma_i^\# \end{matrix} \right\}$ , where  $1 \ll \gamma_i^\# \ll n_i$ . Applying the Jordan limit  $\left\{ \begin{matrix} n \\ a \end{matrix} \right\} \approx a^n / a!$  as  $n \rightarrow \infty$  [25] then gives:

$$\begin{aligned} H_{D:I(g)} &= - \sum_{i=1}^s p_i \ln \frac{p_i}{\gamma_i^\#} - \lim_{N \rightarrow \infty} \frac{1}{N} (\ln k! + \ln \gamma_i^\#!) \\ &= - \sum_{i=1}^s p_i \ln \frac{p_i}{\gamma_i^\#} \quad \text{for } k \nrightarrow \infty, \gamma_i^\# \nrightarrow \infty \end{aligned} \quad (19)$$

This closely resembles the degenerate Maxwell-Boltzmann entropy  $H_{MB} = - \sum_{i=1}^s p_i \ln(p_i/g_i)$ , where  $g_i$  is the degeneracy of state  $i$  [15, 16, 17]. However, as shown in §III, this form does not reflect the behaviour of this statistic when unfilled states or sub-states are present, for which  $H_{D:I(g)}^\#$  (18) must be used.

#### V. CONCLUSIONS

The statistical weight for the allocation of distinguishable entities to indistinguishable states is derived herein, for both non-degenerate or equally degenerate states. The weight is obtained as a function of a multinomial coefficient, a set of coding parameters, and (for the degenerate case) a set of Stirling numbers of the second kind or of incomplete Bell numbers. Using Boltzmann's combinatorial definition (the ‘‘Boltzmann principle’’), the non-asymptotic entropy functions are then obtained. For fully filled states, the non-degenerate and degenerate entropies converge respectively to the Shannon and degenerate Maxwell-Boltzmann functions, but not otherwise.

This study illustrates the importance of the combinatorial definition of entropy, for which the maximum entropy position (‘‘MaxEnt’’) gives the most-probable (‘‘MaxProb’’) realization of the system (or, in the non-asymptotic case, a distribution close to the maximum). For systems which follow the  $D : I$  statistic, blind application of MaxEnt based on the Shannon entropy (3) will give the most probable realization only in special circumstances. Given the long history of Bose-Einstein and Fermi-Dirac statistics in physics, for the allocation of indistinguishable entities to distinguishable states [10, 11, 12, 13, 14, 15, 16, 17], it is surprising that the entropy functions for the opposite occupancy problem do not appear to have been examined previously.

#### Acknowledgments

The author thanks the organisers of the CTNEXT07 conference, Catania, Italy, at which this work was presented; the European Commission for financial support by a Marie Curie Fellowship under FP6; and The University of New South Wales for sabbatical leave.

$N$	$s$	MaxProb realization $[n_i]$	Maxima	$\mathbb{W}$ (each)	$\mathbb{P} = \frac{\mathbb{W}}{\sum \mathbb{W}}$ (each)
1	1	[1]	1	1	1
2	2	[1, 1]	1	2	0.5
3	3	[1, 1, 1]	1	6	0.222222
4	4	[1, 1, 1, 1]	1	24	0.093750
5	5	[1, 1, 1, 1, 1]	1	120	0.038400
10	10	[1, 1, 1, 1, 1, 1, 1, 1, 1, 1]	1	3.63E+06	3.629E-04
20	20	[1, ..., 1]	1	2.43E+18	2.320E-08
30	30	[1, ..., 1]	1	2.65E+32	1.288E-12
40	40	[1, ..., 1]	1	8.16E+47	6.749E-17
50	50	[1, ..., 1]	1	3.04E+64	3.424E-21
1	3	[1, 0, 0], [0, 1, 0], [0, 0, 1]	3	1	0.333333
2	3	[1, 1, 0], [1, 0, 1], [0, 1, 1]	3	2	0.222222
3	3	[1, 1, 1]	1	6	0.222222
4	3	[1, 1, 2], [1, 2, 1], [2, 1, 1]	3	12	0.148148
5	3	[1, 2, 2], [2, 1, 2], [2, 2, 1]	3	30	0.123457
10	3	[3, 3, 4], [3, 4, 3], [4, 3, 3]	3	4200	0.071127
20	3	[6, 7, 7], [7, 6, 7], [7, 7, 6]	3	1.33E+08	0.038151
30	3	[10, 10, 10]	1	5.55E+12	0.026961
40	3	[13, 13, 14], [13, 14, 13], [14, 13, 13]	3	2.41E+17	0.019853
50	3	[16, 17, 17], [17, 16, 17], [17, 17, 16]	3	1.15E+22	0.016005

TABLE II: Most probable realizations for the multinomial statistic, using  $\mathbb{W} = \mathbb{W}_{mult}$  and  $\sum \mathbb{W} = s^N$ .

$N$	$s$	MaxProb realization $\{n_i\}$	Maxima	$\mathbb{W}$ (each)	$\mathbb{P} = \frac{\mathbb{W}}{\sum \mathbb{W}}$ (each)
1	1	{1}	1	1	1
2	2	{1, 1}, {2, 0}	2	1	0.5
3	3	{2, 1, 0}	1	3	0.6
4	4	{2, 1, 1, 0}	1	6	0.4
5	5	{2, 2, 1, 0, 0}	1	15	0.288462
10	10	{3, 2, 2, 1, 1, 0, 0, 0, 0}, {3, 2, 2, 2, 1, 0, 0, 0, 0}, {3, 3, 2, 1, 1, 0, 0, 0, 0}, {4, 3, 2, 1, 0, 0, 0, 0, 0}	4	12600	0.108644
20	20	{4, 3, 3, 2, 2, 2, 1, 1, 0, ..., 0}, {4, 4, 3, 3, 2, 2, 1, 1, 0, ..., 0}	2	1.83E+12	0.035443
30	30	{5, 4, 4, 3, 3, 3, 2, 2, 1, 1, 0, ..., 0}	1	1.54E+22	0.018214
40	40	{5, 4, 4, 4, 3, 3, 3, 3, 2, 2, 2, 1, 1, 1, 0, ..., 0}	1	1.14E+33	0.007265
50	50	{6, 5, 5, 4, 4, 4, 3, 3, 3, 3, 2, 2, 2, 1, 1, 0, ..., 0}	1	7.40E+44	0.003986
1	3	{1, 0, 0}	1	1	1
2	3	{1, 1, 0}, {2, 0, 0}	2	1	0.5
3	3	{2, 1, 0}	1	3	0.6
4	3	{2, 1, 1}	1	6	0.428571
5	3	{2, 2, 1}	1	15	0.365854
10	3	{5, 3, 2}	1	2520	0.256046
20	3	{8, 7, 5}	1	9.98E+07	0.171680
30	3	{11, 10, 9}	1	5.05E+12	0.147059
40	3	{15, 13, 12}	1	2.09E+17	0.103236
50	3	{18, 17, 15}	1	1.02E+22	0.085360

TABLE III: Most probable realizations for the D:I statistic, using  $\mathbb{W} = \mathbb{W}_{D:I}$  (9) and  $\sum \mathbb{W} = B(N, s)$  (12).

- 
- [1] L. Boltzmann, Wien. Ber. 76 (1877) 373; English transl.: J. Le Roux (2002) <http://www.essi.fr/~leroux/>.
  - [2] M. Planck, Annalen der Physik 4 (1901) 553.
  - [3] I. Vincze, Progress in Statistics (European Meeting of Statisticians, Budapest, Hungary, 1972), 2 (1974) 869-895.
  - [4] M. Grendár, Jr. M. Grendár, What is the question that MaxEnt answers? A probabilistic interpretation, in A. Mohammad-Djafari (ed.), Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt 2000), AIP, Melville, 2001, 83-94.
  - [5] R. K. Niven, Combinatorial information theory: I. Philosophical basis of cross-entropy and entropy, *arXiv:0512017*, 2005-2007.
  - [6] R.K. Niven, Origins of the combinatorial basis of entropy, MaxEnt2007, 8-13 July 2007, Saratoga Springs, NY, in press, *arXiv:0708.1861*.
  - [7] J. Stirling, Methodus Differentialis: Sive Tractatus de Summatione et Interpolatione Serierum Infinitarum, Gul. Bowyer, London, 1730.
  - [8] I. N. Sanov, Mat. Sbornik, 42 (1957) 1144 (Russian).
  - [9] C.E. Shannon, Bell Sys. Tech. J. 27 (1948) 379; 623.
  - [10] S.N. Bose, Z. Phys. 26 (1924) 178.
  - [11] A. Einstein, Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl (1924) 261.
  - [12] A. Einstein, Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl (1925) 3.
  - [13] E. Fermi, Z. Phys. 36 (1926) 902.
  - [14] P.A.M. Dirac, Proc. Roy. Soc. 112 (1926) 661.
  - [15] R.C. Tolman, The Principles of Statistical Mechanics, Oxford Univ. Press, London, 1938.
  - [16] N. Davidson, Statistical Mechanics, McGraw-Hill, NY, 1962.
  - [17] R.K. Niven, Phys. Lett. A 342(4) (2005) 286.
  - [18] M. Abramowitz, I.A. Stegun, Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables, U.S. Govt Printing Office, Washington D.C, 1965.
  - [19] L. Comtet, Advanced Combinatorics, rev. ed., D. Reidel, Dordrecht.
  - [20] C.A. Charalambides, Enumerative Combinatorics, Chapman & Hall / CRC, Boca Raton.
  - [21] E.T. Jaynes, Phys. Rev. 106 (1957) 620.
  - [22] N.L. Johnson, S. Kotz, Urn Models and Their Application, John Wiley, NY, §1.3.4.
  - [23] K.-T. Fang, in S. Kotz, N.L. Johnson (eds), Encyclopedia of Statistical Sciences, vol. 6, John Wiley, NY, 1985, 402-406.
  - [24] D. Zwillinger, CRC Standard Mathematical Tables and Formulae, Chapman & Hall / CRC Press, Boca Raton, FL, 2003.
  - [25] C. Jordan, Calculus of Finite Differences, Chelsea Publ., NY, 1947.